RankTower: A Synergistic Framework for Enhancing Two-Tower Pre-Ranking Model

YaChen Yan, Liubo Li

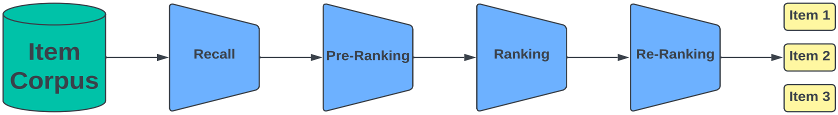
August 16, 2024

Summary

Introduction

The Architecture of Cascade Ranking System

A typical cascade ranking system consists of multiple sequential stages, including recall, pre-ranking, ranking, and re-ranking stages.



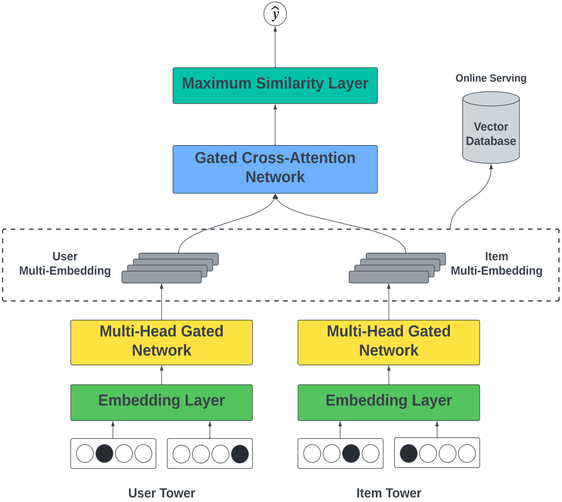
Pre-ranking models are required to score a larger number of candidate items for each user and demonstrate higher inference efficiency than ranking models.

Proposed model: RankTower

- Designed a novel neural network architecture for effectively and efficiently learning user-item interactions
- Integrated full-stage sampling strategy and hybrid loss function to learn ordering dynamics within a cascade ranking system
- Extensive experiments demonstrates proposed model significantly outperforms state-of-the-art pre-ranking models on public datasets

RankTower

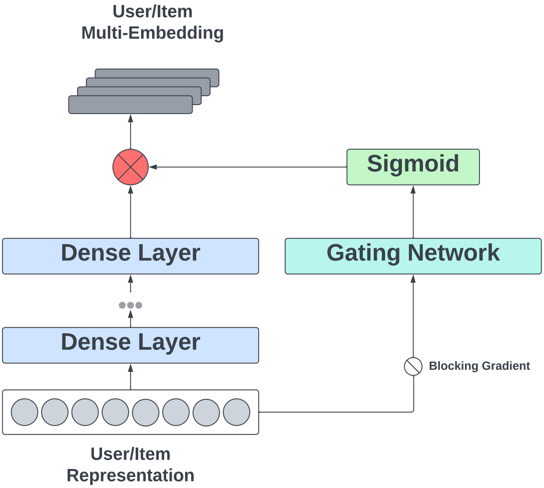
The Architecture of RankTower



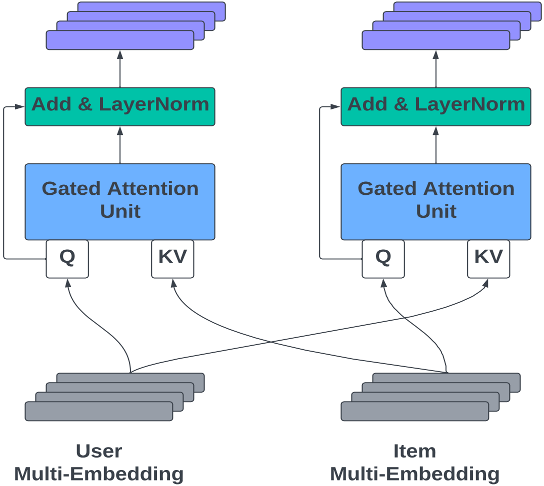
The Architecture of RankTower Cont.

- Three main components
  - Multi-Head Gated Network
  - Gated Cross-Attention Network
  - Maximum Similarity Layer
- Following user-item decoupling paradigm for efficient online serving
- Pre-computing and caching user and item representations

Multi-Head Gated Network



Gated Cross-Attention Network



Maximum Similarity Layer

The Maximum Similarity Layer computes the final probability prediction based on the user and item attended embeddings.

$$s = (\sum_{p=1}^{H_u} \operatorname*{Max}_{q \in \{1, \cdots, H_i\}} COSINE(\mathcal{E}_u^p, \mathcal{E}_i^q))/\tau \quad (1)$$

where p and q are the sub-space indexes of user-attended embedding and item-attended embedding, respectively, and τ is the learnable temperature scalar for re-scaling the cosine similarity.

Model Optimization

- Sampling Strategy
- Label Aggregation
- Hybrid Loss Function

Full-Stage Sampling Strategy

The RankTower model is trained using user-level listwise samples containing multiple positive and netagive items. The training samples for each user are sourced from various stages of the cascade ranking system.

- Impression Samples
- Candidate Samples
- Random Samples

Label Aggregation

- Hard labels: we aggregate labels according to their orders of importance. In an e-commerce context, one might establish a relative preference order based on the depth of user feedback, such as $Purchase \geq Add\ to\ Cart \geq Click$.
- Soft labels: we use the ranking objective function as aggregation function.

Hybrid Loss Function

- Distillation Loss (Softmax)

$$\mathcal{L}_{\text{Distillation}}(z, p) = -\sum_{i \in \mathcal{D}_l} p_i \log \frac{\exp(z_i)}{\sum_{j \in \mathcal{D}_l} \exp(z_j)}$$

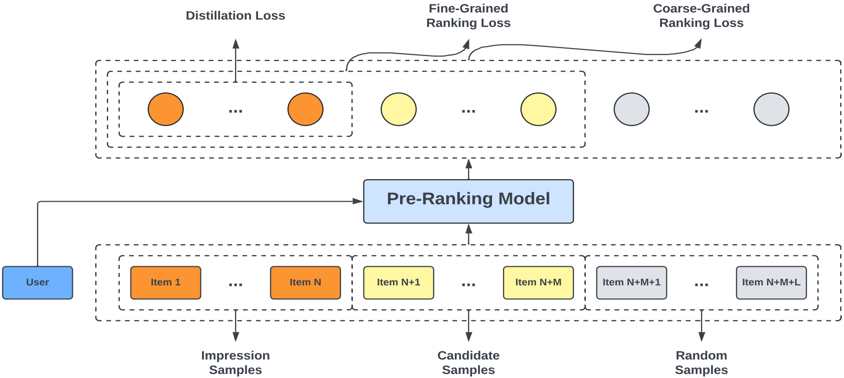
- Fine-Grained Ranking Loss (SoftSort)

$$\mathcal{L}_{\text{Sorting}}(z, y) = -\mathbf{tr}\Big(\mathbf{J}_n\big(\mathbf{SoftSort}_\tau^d(y) \circ \log \mathbf{SoftSort}_\tau^d(z)\big)\Big)$$

- Coarse-Grained Ranking Loss (Adaptive Margin Rankmax)

$$\mathcal{L}_{Rankmax}(z, y) = \sum_{j: y_j > 0} \log \sum_{i=1}^{n} \big(z_i - z_j + 1\big)_+$$

Hybrid Loss Function Cont.



Experiment

Experimental Setup

- Datasets: Alimama, Taobao, KuaiRand
- Evaluation metrics: Recall@K, NDCG@K
- Competing models: LR, Two-Tower, DAT, COLD, IntTower, ARF

Model Performance Comparison

	Alimama		Taobao		KuaiRand	
Model	Recall@K	NDCG@K	Recall@K	NDCG@K	Recall@K	NDCG@K
LR	0.4802	0.3237	0.4792	0.2685	0.6713	0.5027
Two-Tower	0.5123	0.3428	0.5019	0.2921	0.6902	0.5258
DAT	0.5161	0.3472	0.5089	0.3013	0.6955	0.5312
COLD	0.5210	0.3518	0.5123	0.3070	0.7011	0.5349
IntTower	0.5215	0.3519	0.5101	0.3051	0.6960	0.5309
ARF	0.5318	0.3655	0.5215	0.3117	0.7096	0.5497
RankTower	**0.5462**	**0.3794**	**0.5301**	**0.3223**	**0.7182**	**0.5551**

Experiment Results for Different Sampling Strategies

	Recall@K	NDCG@K
Full-Stage Sampling	0.7182	0.5551
w/o random samples	0.7125	0.5437
w/o candidate samples	0.7040	0.5401
w/o candidate & random samples	0.6981	0.5323

Experiment Results for Different Ranking Losses

	Recall@K	NDCG@K
Hybrid Loss	0.7182	0.5551
Sorting	0.7128	0.5516
AM-Rankmax	0.7132	0.5507
Rankmax	0.7105	0.5492
Softmax	0.7109	0.5498
ApproxNDCG	0.7006	0.5436
RankNet	0.7072	0.5452

Experiment Results for Different Distillation Losses

	Recall@K	NDCG@K
Distillation (Softmax)	0.7182	0.5551
Distillation (Weighted Logloss)	0.7130	0.5519
Distillation (Pairwise Logloss)	0.7071	0.5432
No Distillation	0.7108	0.5495

Conclusions

Conclusions

- RankTower architecture is effectively and efficiently capturing bi-directional latent user-item interactions by integrating Multi-Head Gated Network, Gated Cross-Attention Network, and Maximum Similarity Layer.
- The integrated full-stage sampling strategy and hybrid loss function ensure the ranking consistency within cascade ranking system

Thank You!